Elementary Linear Fitting Theory

Rajendran Raja
18-Oct-2004

We outline the elements of linear least squares fitting theory and show how to obtain errors on fitted parameters. The theory is in fact general for all cases where the dependence of the predicted values on fitted parameters can be linearized, i.e. the likelihood function is Gaussian near the maximum likelihood point.. We illustrate the main ideas using track fitting as a concrete example.

Notation

Let u_i^p ,i=1,n denote the predicted co-ordinates of the track at wire plane i at z co-ordinate z_i. The orientation of the wire plane with respect to the x- axis is given by the unit vector b, which is perpendicular to the z- axis. Let the track fitted parameters be denoted by a_k (k=1,p). In three dimensions, the number of parameters to describe a track are the x,y co-ordinates at a particular value of z, the momentum and the direction of the track at the x,y point. This makes for a total of 5 parameters, i.e p=5. If the momentum is not being fit, then only four parameters are needed. In this case p=4.

The error matrix E of the measured quantities u_i is defined by

$$E_{ij} = \left\langle (u_i - \langle u_i \rangle)(u_j - \langle u_j \rangle) \right\rangle = \langle u_i u_j \rangle - \langle u_i \rangle \langle u_j \rangle \qquad (1)$$

where the brackets <> are meant to denote the average over many events. If the quantities i, and j are un-correlated for $i \neq j$, then E is a diagonal matrix. For tracks, this is clearly the case, since the measurements of the various planes are un-correlated. The diagonal elements of the matrix E are the variances of the quantities u_i, denoted by σ^2_{i}.

Let us denote by the vector X_i, the quantities u_i-u_i^p. Let the matrix H denote the inverse of the error matrix E. H is known in the jargon as the Hessian matrix. Then the generalized χ^2 (valid in the presence of correlations) of a fit is given by

$$c^2 = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} H_{ij} X_i X_j \qquad (2)$$

Since the predicted values u_i^p are functions of the track fit parameters a_k,k=1,p, then so is χ^2.

Fitting

Fitting is performed by minimizing χ^2 with respect to the parameters $a_{k,}$k=1,p. We linearize the problem by assuming that the predicted values u_i^p are linear functions of the parameters a_k. This is certainly true of straight line fits. i.e

$$u_i^p = l_{ik} a_k$$

where λ_{ik} is a n by p matrix that is not square in general. *Repeated indices imply summing over* (i.e in the above equation, we imply summing over the index k).

Then χ^2 can be expressed as

$$c^2 = H_{ij}(u_i - u_i^p)(u_j - u_j^p) = H_{ij}u_i u_j - 2H_{ij}u_i u_j^p + H_{ij}u_i^p u_j^p \qquad (3)$$

where we have used the fact that H is a symmetric matrix, i.e $H_{ij}=H_{ji}$.

At the minimum of χ^2,

$$\frac{\partial c^2}{\partial a_k} = -2H_{ij}u_i \frac{\partial u_j^p}{\partial a_k} + 2H_{ij}u_i^p \frac{\partial u_j^p}{\partial a_k} = 0$$

Linearizing,

$$\frac{\partial c^2}{\partial a_k} = -2H_{ij}u_i l_{jk} + 2H_{ij}u_i^p l_{jk} = 0$$

so at minimum,

$$H_{ij} l_{jk} u_i^p = H_{ij} l_{jk} u_i$$

substituting for u_i^p.

$$H_{ij} l_{jk} l_{il} a_l = H_{ij} l_{jk} u_i$$

The above set of equations (there are p of them) can be abbreviated in matrix form by

$$M_{kl} a_l = N_k$$

where the square (p by p) matrix $M_{kl}=H_{ij}\lambda_{jk}\lambda_{il}$ and the (1 by p) row vector $N_k = H_{ij}\lambda_{jk}u_{i.}$

This matrix equation can be inverted yielding the fitted parameters $a_{k.}$

$$a_k = \left(M^{-1}\right)_{kl} N_l$$

Errors of the fitted parameters

Just as the measured quantities u_i possess an error matrix E and its inverse H (these are n by n matrices), the fitted quantities a_k also possess an error matrix ε and its inverse denoted by η. These are (p by p) matrices. Let us denote the minimum χ^2 as χ^2_{min} and the fitted parameters as $a^*_{m,}$m=1,p. We need to ask how the χ^2 changes as we change the parameters by small increments δa_m away from the minimum. As we change the parameters, the predicted values will change by δu_i^p and the χ^2 change can be written (using equation (3))

$$dc^2 = H_{ij} du_i^p du_j^p = H_{ij} \frac{\partial u_i^p}{\partial a_k} \frac{\partial u_j^p}{\partial a_l} da_k da_l = H_{ij}\, l_{ik}\, l_{jl} da_k da_l = M_{kl} da_k da_l$$

Just as in measurement space, the minimization of χ^2 implies maximizing a Gaussian likelihood function, with an error matrix given by E, as in equation (1 and 2), in parameter space, the same likelihood function is expressed by the error matrix ε of the parameters and its inverse Hessian η. Thus, using equation (4),

$$dc^2 = h_{kl} da_k da_l = M_{kl} da_k da_l$$

Since this is true for arbitrary changes of parameters δa_k, this implies η=M, leading to

$$e = M^{-1}$$

It should be noted that χ^2 is distributed as a Γ variate with n-p degrees of freedom. To show this is straight forward, but beyond the scope of this write-up.

Equations for straight lines in 3 dimensions

The above theory (called linearized least squares fitting) assumes that near the χ^2 minimum, the χ^2 as a function of the parameters is parabolic. Most problems can be approximated in this fashion near the minimum to second order. The first order terms in a Taylor expansion are zero (since it is a minimum) and the second order terms dominate and the third order terms can be neglected. This is how the MINUIT program (subprogram MIGRAD) works.

We can use the above theory for fitting helices as well as straight lines in the beam chambers. To do this, all one needs to do is to work out u_i^p as a function of the straight line parameters in three dimensions. For the beam chambers, we will work in cylindrical co-ordinates with the z-axis being the axis of symmetry. Then the equation for a straight line in 3D is

$$\vec{r} = \vec{c} + \vec{m}z$$

where $\vec{r}$ is the vector denoting a point on the straight line at z co-ordinate z and $\vec{c}$ is the value of $\vec{r}$
when z=0. $\vec{m}$ is the slope vector whose x,y,z components are given by (tanθ cos φ, tan θ sin φ, 1).
θ is the angle wrt z axis and φ is the angle of the projection of the straight line in the x-y plane wrt x-axis.

The wire planes are specified by their z co-ordinate z_i, i=1,n and their b vectors $\mathbf{b_i}$ such that the predicted co-ordinate u_i^p is given by

$$u_i^p = \vec{r}_i . \vec{b}_i \qquad (5)$$

The vector $\mathbf{b_i}$ has components ($\cos\beta_i$, $\sin\beta_i$, 0) where β_i is the angle of the wires w.r.t to the y –axis. An x-measuring chamber has wires running along the y –axis.

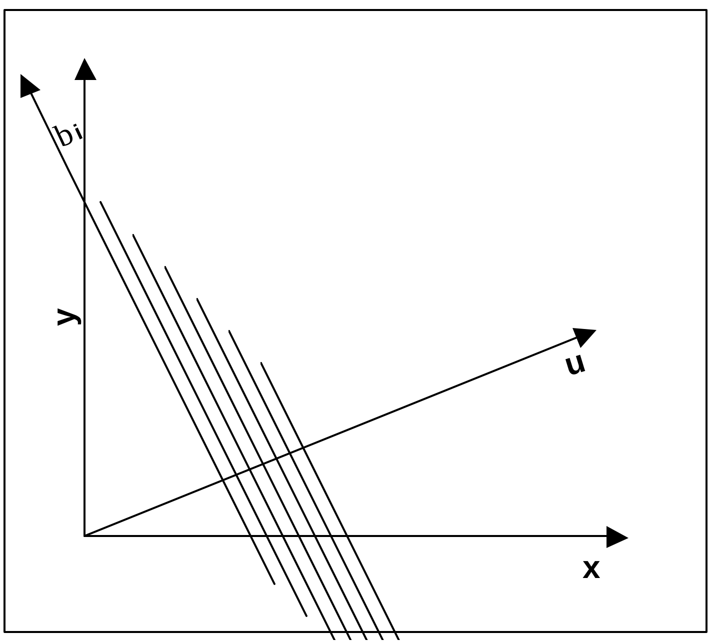

Figure 1 Chamber geometry with wire directions shown. z axis is out of the paper. The sign of the angle b_i is positive counter-clockwise from the y axis.

With these conventions, the parameters of the track (p=4) can be taken to be (m_x,m_y,c_x and c_y). With these conventions, equation 5 reduces to

$$u_i^p = c_x b_{ix} + c_y b_{iy} + m_x b_{ix} z_i + m_y b_{iy} z_i$$

There is no implied sum over i in the above equation or the ones below.

The functions λ_{ik} can thus be trivially determined by differentiation with the convention k=1,4 implying in turn (c_x, c_y, m_x and m_y).

$$
\begin{aligned}
l_{i1} &= b_{ix} && ;\text{ for } c_x \\
l_{i2} &= b_{iy} && ;\text{ for } c_y \\
l_{i3} &= b_{ix} z_i && ;\text{ for } m_x \\
l_{i4} &= b_{iy} z_i && ;\text{ for } m_y
\end{aligned}
$$

We can now proceed with the fit. This involves inverting a 4 by 4 matrix M. The fitted parameters are correlated since the wire planes are neither pure x nor y measuring planes.